

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
29 March 2001 (29.03.2001)

PCT

(10) International Publication Number  
**WO 01/21782 A2**

(51) International Patent Classification<sup>7</sup>: C12N 15/00

(21) International Application Number: PCT/US00/26052

(22) International Filing Date:  
22 September 2000 (22.09.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/156,071 24 September 1999 (24.09.1999) US

(71) Applicant: CALGENE LLC [US/US]; 1920 Fifth Street,  
Davis, CA 95616 (US).

(72) Inventor: CHAUDHURI, Sumita; 1125 Juniper Place,  
Davis, CA 95616 (US).

(74) Agent: MCBRIDE, Thomas, P.; Monsanto Company,  
800 North Lindbergh Boulevard, St. Louis, MO 63167  
(US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published:  
— Without international search report and to be republished upon receipt of that report.

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

WO 01/21782 A2

(54) Title: ENHANCER ELEMENTS FOR INCREASED TRANSLATION IN PLANT PLASTIDS

(57) Abstract: Provided are methods for increasing the production of protein in a plant cell by transforming plastids of plant cells with a construct comprising a translation initiation site, DNA sequence of interest and a transcription termination region, and growing plant cells comprising the transformed plastids under conditions wherein the DNA encoding sequence is transcribed in the plastid.

**ENHANCER ELEMENTS FOR INCREASED TRANSLATION  
IN PLANT PLASTIDS**

5

**INTRODUCTION**

Technical Field

This invention relates to the application of genetic engineering techniques to plants. Specifically, the invention relates to compositions and methods for enhancing expression of proteins in plant plastids.

Background

The plastids of higher plants are an attractive target for genetic engineering. Plant plastids (chloroplasts, amyloplasts, elaioplasts, etioplasts, chromoplasts, etc.) are the major biosynthetic centers that, in addition to photosynthesis, are responsible for production of industrially important compounds such as amino acids, complex carbohydrates, fatty acids, and pigments. Plastids are derived from a common precursor known as a proplastid, and thus the plastids present in a given plant species all have the same genetic content. Plant cells contain 500-10,000 copies of a small 120-160 kilobase circular genome, each molecule of which has a large (approximately 25kb) inverted repeat. Thus, it is possible to engineer plant cells to contain up to 20,000 copies of a particular gene of interest, which potentially can result in very high levels of foreign gene expression. In addition, plastids of most

plants are maternally inherited. Consequently, unlike heterologous genes expressed in the nucleus, heterologous genes expressed in plastids are not pollen disseminated, therefore, a trait introduced into a plant plastid will not be transmitted to wild-type relatives.

There remains a need for improved regulatory elements for expression of genes in a plant plastid. To date, the expression signals used routinely for plastid transgene expression derive from endogenous plastid genes. The plastid expression signals are typically derived from promoter regions of highly expressed plastid genes such as the promoter regions from the 16S ribosomal RNA operon (*Prrn*), *psbA* gene (*PpsbA*) or the *rbcL* gene (*PrbcL*). The *psbA* and *rbcL* genes are highly transcribed, but their translation is controlled by tissue-specific and light-regulated factors which limits their usefulness. In the case of *Prrn*, a synthetic ribosome binding site (RBS) patterned after the plastid *rbcL* gene leader has been typically used to direct translation. DNA sequences expressed from this *Prrn*/RBS accumulate high levels of the transcript; however, the mRNA is translated inefficiently resulting in poor gene expression.

A totally heterologous expression system has been used to express plastid genes (USPN 5,576,198, the entirety of which is incorporated herein by reference). This system is a two-component system. The first component is a plastid transgene driven by a T7 bacteriophage gene 10 promoter/leader sequence. The second component is a nuclear gene encoding the T7 Polymerase that is targeted to the

plastid compartment. The limitation of this system is the need to create nuclear transformed lines that express the T7 Polymerase in preferred ways.

Plastids of higher plants present an attractive target  
5 for genetic engineering. As mentioned above, plastids of higher plants are maternally inherited. This offers an advantage for genetic engineering of plants for tolerance or resistance to natural or chemical conditions, such as herbicide tolerance, as these traits will not be transmitted  
10 to wild-type relatives. In addition, the high level of foreign gene expression is attractive for engineered traits such as the production of pharmaceutically and industrially important proteins.

Expression of nucleic acid sequences encoding for  
15 enzymes providing for herbicide tolerance as well as pharmaceutical proteins and various industrial applications from plant plastid genome offers an attractive alternative to expression from the plant nuclear genome.

## 20 Relevant Literature

McBride *et al.* U.S. Patent Number 5,576,198 and McBride *et al.* (1994) *Proc Natl Acad Sci* 91:7301-7305 reports the plastid expression system based on a two-component system utilizing a nuclearly encoded T7 polymerase targeted to the  
25 plastid that activates a transgene controlled by the T7 bacteriophage gene 10 promoter. Svab *et al.* (1990) *Proc Natl Acad Sci* 87:8526-8530 reports the standard chloroplast transforamtion methods. Svab *et al.* (1993) *Proc Natl Acad Sci* 90:913-917 reports the use of the *aadA* gene for use in

selection of transplastomic plants on spectinomycin and streptomycin, as well as integration sequences. Zoubenko et al. (1994) *Nuc Acid Res* 22:3819-3824 reports the construct of vectors for use in plastid transformation.

- 5        Barry et al., U.S. Patent Number 5,627,061 describes the cloning of EPSPS nucleotide sequences from several sources including *Agrobacterium* strain CP4 and methods for producing glyphosate tolerant plants. Kishore and Shah, *Ann. Rev. Biochem.* (1988) 57:627-663 reports the
- 10       modification of DNA sequences for EPSPS for the enhancement of glyphosate tolerance. Stalker et al., U.S. Patent 4,810,648 describes the cloning and use of a nucleic acid sequence encoding for the bromoxynil degrading gene, nitrilase.

15

#### SUMMARY OF THE INVENTION

- The present invention provides nucleic acid sequences useful in enhancing expression of a wide variety of genes,
- 20       both eukaryotic and prokaryotic, in plant plastids.
- Furthermore, plastid expression constructs are provided that are useful for genetic engineering of plant cells and that provide for enhanced expression of DNA sequences of interest in plant cell plastids. The transformed plastids should be
- 25       metabolically active plastids and are preferably maintained at a high copy number in the plant tissue of interest, most preferably the chloroplasts found in green plant tissues, such as leaves or cotyledons.

The plastid expression constructs for use in this invention generally include a promoter region capable of providing for enhanced expression of a DNA sequence in plastids, a DNA sequence of interest, and a transcription  
5 termination region capable of terminating transcription in a plant plastid.

The plastid promoter region of the present invention is preferably linked to a ribosome binding site comprising downstream box sequences which provide for enhanced  
10 translation of mRNA transcripts in a plant plastid.

The plastid expression construct of this invention is preferably linked to a construct having a DNA sequence encoding a selectable marker that can be expressed in a plant plastid. Expression of the selectable marker allows  
15 the identification of plant cells comprising a plastid expressing the marker.

In a preferred embodiment, vectors for transfer of the construct into a plant cell include means for inserting the expression and selection constructs into the plastid genome.  
20 This preferably comprises regions of homology to the target plastid genome, which flank the constructs.

The constructs of the present invention preferably comprise a promoter sequence linked to sequence elements that promote ribosome binding capable of enhancing the  
25 translation of mRNA transcripts in the plant plastid. In particular, the enhancer sequence contains sequence complementary to a region in the 16S rRNA, referred to as downstream box sequences (also referred to herein as DBs).

Of particular interest in the present invention is the high level of expression of nucleic acid sequences in plant plastids. The present invention therefore also includes methods for enhanced expression of DNA sequences in plant  
5 plastids.

Plant cell plastids containing the constructs are also contemplated in the invention, as are plants, plant seeds, plant cells or progeny thereof containing plastids comprising the construct.

10 Thus, the present invention relates to a chimeric gene containing a DNA sequence of interest, a plant plastid expression vector containing a promoter operably linked to a ribosome binding site containing downstream box sequences capable of enhanced expression in a plant plastid operably  
15 linked to a DNA sequence of interest, a plant transformation vector having inserted therein a DNA sequence of interest expressed from a plastid promoter linked to a ribosome binding site containing downstream box sequences, plant cells transformed using such vectors and plants regenerated  
20 therefrom which exhibit a substantial degree of expression of nucleic acid sequences and proteins and methods for producing such plants and such plants.

25

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 provides a schematic representation of the plasmid pCGN6115.

Figure 2 provides sequence of the ribosome binding site containing the wild-type Shine-Dalgarno sequence, initiation

codon and wild-type downstream box sequence from the T7 bacteriophage gene 10. SEQ ID NO:1

Figure 3 provides sequence of the ribosome binding site containing the mutant Shine-Dalgarno sequence, initiation  
5 codon and wild-type downstream box sequence from the T7 bacteriophage gene 10. SEQ ID NO:2

Figure 4 provides sequence of the ribosome binding site containing the wild-type Shine-Dalgarno sequence, initiation codon and m1 mutant (15/15) downstream box sequence from the  
10 T7 bacteriophage gene 10. SEQ ID NO:3

Figure 5 provides sequence of the ribosome binding site containing the mutant Shine-Dalgarno sequence, initiation codon and m1 mutant (15/15) downstream box sequence from the T7 bacteriophage gene 10. SEQ ID NO:4

15 Figure 6 provides sequence of the ribosome binding site containing the wild-type Shine-Dalgarno sequence, initiation codon and the m2 mutant (11/15) downstream box sequence from the T7 bacteriophage gene 10. SEQ ID NO:5

Figure 7 provides sequence of the ribosome binding site  
20 containing the wild-type Shine-Dalgarno sequence, initiation codon and the m3 mutant (0/15) downstream box sequence from the T7 bacteriophage gene 10. SEQ ID NO:6

Figure 8 provides a schematic representation of the plastid transformation vector comprising the T7 promoter, E1  
25 cellulase, and terminator sequences for *psbA* and T7. The Lg10 fragment comprises various Shine-Dalgarno and downstream box sequences. The various combinations are schematically depicted.



Figure 9 provides complementary base pairings between the Shine-Dalgarno and downstream box sequences of the T7 gene 10 ribosome binding site and the bacterial and plastidial 16S rRNA sequences.

5

#### DETAILED DESCRIPTION OF THE INVENTION

In accordance with the subject invention, plastid expression constructs are provided that generally comprise a promoter functional in a plant plastid, a translation  
10 initiation site and a sequence downstream of the initiation site capable of providing for enhanced expression, a DNA sequence encoding for a gene of interest, and a transcription termination region capable of terminating transcription in a plant plastid. These elements are  
15 provided as operably joined components in the 5' to 3' direction of transcription.

Of particular interest in the present invention is the use of sequences downstream of the translation initiation site, referred to as downstream box (DB) sequences, in the  
20 preparation of constructs for plant transformation.

Also of particular interest in the present invention is the use of the plastid expression constructs containing the RBS sequences of the present invention to direct the high level transcription and translation (expression) of nucleic  
25 acid sequences. Such plastid expression constructs find use in directing high level expression of DNA sequences encoding for enzymes involved in a variety of plant genetic engineering applications.

Of more particular interest in the present invention is the use of the plastid expression constructs containing the RBS sequences of the present invention to direct the high level translation of transcribed messenger RNA. Such  
5 increases in translation provides for the high level of accumulation of proteins encoded for by DNA sequences of interest expressed from the expression constructs of the present invention.

DNA sequence and biochemical data reveal a similarity  
10 of the plastid organelle's transcriptional and translational machineries and initiation signals to those found in prokaryotic systems. In fact, plastid-derived promoter sequences have been reported to direct expression of reporter genes in prokaryotic cells. In addition, plastid  
15 genes are often organized into polycistronic operons as they are in prokaryotes.

Despite the apparent similarities between plastids and prokaryotes, there exist fundamental differences in the methods used to control gene expression in plastids and  
20 prokaryotes. As opposed to the transcriptional control mechanisms typically observed in prokaryotes, plastid gene expression is controlled predominantly at the level of translation and mRNA stability by trans-acting nuclear encoded proteins.

25 Translation is a multi-stage process beginning with the binding of messenger RNA (mRNA) to ribosomes. Starting at the translation start codon, the mRNA codons are read sequentially as the ribosomes move along the mRNA molecule. The specified amino acids are then sequentially added to the

growing polypeptide chain to yield the protein or polypeptide encoded in the mRNA.

As mentioned, the first step in the translation process is the binding of the mRNA molecule to the ribosome. The nature of this interaction (i.e., binding) has been only partially elucidated. Analysis of RNase-resistant oligonucleotides isolated from bacterial translation initiation complexes indicate that a RNA fragment approximately 30 to 40 nucleotides in length comprises the initial ribosome binding site (RBS). Thus, a RBS is herein understood to comprise a sequence of mRNA surrounding the translation start codon, both upstream and downstream, that is responsible for the binding of the ribosome and for initiation of translation.

Recently, ribosome binding sites have been identified capable of directing enhanced translation in prokaryotes. For example, a ribosome binding site derived from the T7 bacteriophage gene 10 leader, G10L (USPN 5,232,840, the entirety of which is incorporated herein by reference), has been identified that enhances expression of nucleic acid sequences in prokaryotes.

Recent studies of ribosome binding sites in prokaryotes have identified specific sequences in the RBS as important in the translation of mRNA. In *E. coli*, the RBS comprises the start codon and an upstream sequence known as the Shine-Dalgarno sequence. The Shine-Dalgarno sequence has been shown to be responsible for base pairing with the 3' terminus of the 16srRNA. Furthermore, sequences downstream of the start codon, referred to as the downstream box (DB),

have been shown to be important for the high level translation of the gene product (Sprengart et al. (1990) *Nuc. Acid Res.* 18(7):1719-1723). This sequence is complementary to nucleotides 1471-1482 of the 16srRNA, referred to as the anti-downstream box (also referred to herein as ADB). Additional DB sequences that are complementary to this sequence from the 16srRNA have also been identified from other highly expressed genes in *E. coli* as well as *E. coli* bacteriophages. Thus, as used herein, a downstream box sequence is a sequence having a region complementary to a region of ribosomal RNA that associates with message RNA during initiation of translation and is capable of enhancing translation of a coding region operably coupled thereto.

The skilled artisan will recognize base pairing of complementary sequences of DB/ADB, as well as SD/anti-SD, alternative base-pairing interactions between the mRNA and rRNA are possible. For example, in addition to base pairing between Guanine (G) and Cytosine (C) and Adenine (A) and Uracil (U), base pairing interactions between Guanine and Uracil are also possible.

The downstream box (DB) sequence, a stretch of 12-15 nucleotides located downstream of the initiation codon of certain mRNAs, has been shown to be a translation enhancer in bacteria. Phage T7 gene 10 mRNA constructs have been shown to be better translated when they contain such DB sequences (Sprengart et al., 1990 *Nucleic Acids Research* 18:1719-1723; Sprengart et al., 1996, *EMBO J* 3:665-674). DBs are also found in both heat-shock and cold-shock bacterial

genes and have been implicated in translation induction of CspA, a major cold protein in *E. coli* (Mitta et al., 1997, Mol Microbiology 26(2): 321-335). The distinctive feature of the DB is its complementarity to the region generally  
5 between 1469-1483 of the 16S rRNA (anti-DB). The interaction of the DB with anti-DB, analogous to the interaction of the SD with its complementary sequence in the 16S rRNA region, is proposed to be the mechanism that stimulates and enhances translation. Evidence for this interaction is that  
10 increasing the complementarity between DB and anti-DB correlates with increased expression of protein (Sprengart et al., 1996 EMBO J 3:665-674). Furthermore, it has been shown that 30S ribosomes, which complex with mRNA in the first step of translation initiation, have a high affinity  
15 for DB containing RNA sequences (Etchegaray et al., 1998, Mol Microbiology 27(4) 873-874).

Recently, Sprengart et al. ((1996) EMBO J 15(3):665-674) demonstrated that specific mutations of the T7 gene 10 that increases DB/ADB interactions improve the expression of  
20 DNA sequences with such downstream boxes in *E. coli*. Furthermore, mutated downstream box sequences were capable of protein synthesis in the absence of Shine-Dalgarno sequences. In addition, the downstream box was not functional when its position was shifted upstream of the  
25 initiation codon to the position of the Shine-Dalgarno sequence. It was also demonstrated that the downstream box sequence was capable of providing for translation without the Shine-Dalgarno sequence. Thus, by utilizing a

downstream box sequence, translation is achieved in prokaryotic systems.

As described in more detail in the examples that follow, downstream box sequences from the T7 bacteriophage gene 10 (also referred to herein as T7 gene 10) are used in plant plastid expression constructs employing the T7 polymerase expression system (described in U.S. Patent 5,576,198, the entirety of which is incorporated herein by reference).

Sequence comparisons between the T7 gene 10 downstream box and the tobacco plastid 16S rRNA anti-downstream box sequence reveal an approximately 50% complementarity of the nucleotides (7 out of 15 nucleotides are complementary, or 7/15). The complementarity between the T7 gene 10 DB and the bacterial ADB shows complementarity of 11 out of 15 nucleotides (11/15). Thus, constructs are prepared employing various levels of complementarity between the DB nucleotides and the ADB sequence. In a preferred embodiment of the invention, the downstream box sequence is complementary to between about 7 and about 15 bases of the nucleotide sequence consisting of AGUGAUCGGGACGGC (SEQ ID NO:15), including bases complementary through G-U base-pairing. More particularly, the downstream box sequence is selected from SEQ ID NO:16-18.

The downstream box sequences of the present invention find use in the preparation of DNA constructs. The DNA constructs find use in transformation experiments to generate mature plants containing such DNA constructs in the plant cell plastid. Furthermore, the DNA constructs are

capable of directing the high level expression of nucleic acid sequences in the plant cell plastid.

Furthermore, the downstream box sequences of the present invention find use in the preparation of constructs  
5 in conjunction with additional ribosome binding site elements. Addition of other elements, derived from prokaryotic or eukaryotic sources, may allow for an even greater level of expression in a plant cell plastid. Additional elements include, but are not limited to, Shine-  
10 Dalgarno sequences (native or optimized), spacer sequences, stretches of single nucleotides, AU rich sequences for S1 protein binding (Stern et al. (1997) *Trends in Plant Sci.* 2(8):308-315), epsilon sequences (Olins et al. (1989) *J. Biol. Chem.* 264:16973-16976), additional N-terminal  
15 extensions, and the like.

Shine-Dalgarno sequences are known in prokaryotic expression systems; however, their role in gene expression in plastids has yet to be elucidated. SD sequences comprise a purine rich sequence and are responsible for binding near  
20 the 3' terminus of the 16S rRNA via base-pairing of complementary sequences. In prokaryotic systems, the SD sequence is generally found 3 to 12 nucleotides upstream (i.e. 5') of the start codon, and contains from 4 to 8 nucleotides. The optimal spacing between the start codon  
25 and the SD is 7 to 9 nucleotides (Gold (1988) *Annu. Rev. Biochem.* 57:199-233; and Ringquist et al. (1992) *Mol. Microbiol.* 6:1219-1229).

Addition of sequences to limit the amount of secondary structures in the mRNA may also provide a means for enhanced

expression. Reduction of secondary structures allows for the accessibility of the SD and DB sequences base-pairing with the anti-SD and anti-DB sequences of the 16S rRNA and for ribosome binding.

5        In addition, inclusion of AU sequences upstream of the SD sequence may also provide for enhanced expression (Hirose et al. (1996) *EMBO J.* 15(7):1687-1695). AU-rich sequences are found in the 5'-UTR of several plastidial DNA sequences. Such sequences bind with S1 proteins. The S1 protein may be  
10 involved in recognition of specific RNA structures and allows for ribosomal binding in the absence of a Shine-Dalgarno sequence (Tzareva et al. (1994) *FEBS Let.* 337:189-194; Ringquist et al. (1995) *Biochemistry* 34:3640-3648; Voorma (1996) *Translational Control*, Cold Spring Harbor  
15 Press, pp759-777).

Furthermore, optimal distances between elements may also enhance expression. For example, by addition of an N-terminal extension to a DNA sequence encoding a protein of interest, optimal SD and DB sequences may be utilized, with  
20 spacer sequences between to allow for greater ability to base-pair with the respective complementary sequences of the 16S rRNA. In addition, leader sequences N-terminal to the translation start of greater than about 25 nucleotides may provide for a more stable RNA message, thus leading to a  
25 higher level of expression in a plant cell plastid.

As described in more detail in the examples herein, plastid transformation constructs are prepared comprising the downstream box sequences of the present invention in collaboration with the T7 gene 10 leader sequence comprising



a Shine-Dalgarno sequence to provide for high level expression of E1 cellulase in a plant cell plastid. Increases of at least about 4- to 6-fold are obtained from plants containing plastid containing expression constructs  
5 employing the RBS sequences of the present invention over levels obtained from plants expressing cellulase sequences from expression constructs employing the G10L RBS sequence lacking the DB sequences.

In addition, constructs are prepared that employ an N-  
10 terminal extension containing a synthetic ribosome binding site, comprising an AU rich sequence from the *psbA* RBS, an optimized 8 base pair Shine-Dalgarno sequence, and a start codon. These sequences are optimized for enhancing expression from constructs employing such synthetic RBS  
15 sequences.

The skilled artisan will recognize that the translational enhancer elements of the present invention may be used with various promoter and DNA sequences of interest. The promoter may be derived from the native promoter from  
20 which the ribosome binding site is derived. Alternatively, the RBS sequences may be employed with heterologous promoter sequences.

Furthermore, the RBS sequences employed in the constructs of the present invention may be derived from  
25 various sources. For example, the elements may be derived from prokaryotic sources, such as bacterial systems, as well as from eukaryotic sources, such as plastidial expression systems. Furthermore, the RBS sequences may be derived from such sources and optimized for enhancing translation in a

plant cell plastid. Such elements find use in the preparation of constructs to direct the expression of DNA sequences of interest in a plant cell plastid.

In developing the constructs, the various fragments  
5 comprising the regulatory regions and open reading frame may be subjected to different processing conditions, such as ligation, restriction enzyme digestion, PCR, *in vitro* mutagenesis, linkers and adapters addition, and the like. Thus, nucleotide transitions, transversions, insertions,  
10 deletions, or the like, may be performed on the DNA that is employed in the regulatory regions or the DNA sequences of interest for expression in the plastids. Methods for restriction digests, Klenow blunt end treatments, ligations, and the like are well known to those in the art and are  
15 described, for example, by Maniatis et al. (in *Molecular cloning: a laboratory manual* (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY).

During the preparation of the constructs, the various fragments of DNA will often be cloned in an appropriate  
20 cloning vector, which allows for amplification of the DNA, modification of the DNA or manipulation of the DNA by joining or removing sequences, linkers, or the like. Preferably, the vectors will be capable of replication to at least a relatively high copy number in *E. coli*. A number of  
25 vectors are readily available for cloning, including such vectors as pBR322, vectors of the pUC series, the M13 series vectors, and pBluescript vectors (Stratagene; La Jolla, CA).

In order to provide a means of selecting the desired plant cells, vectors for plastid transformation typically

contain a construct that provides for expression of a selectable marker gene. Marker genes are plant-expressible DNA sequences that express a polypeptide that resists a natural inhibition by, attenuates, or inactivates a selective substance, i.e., antibiotic, herbicide etc.

Alternatively, a marker gene may provide some other visibly reactive response, i.e., may cause a distinctive appearance or growth pattern relative to plants or plant cells not expressing the selectable marker gene in the presence of some substance, either as applied directly to the plant or plant cells or as present in the plant or plant cell growth media.

In either case, the plants or plant cells containing such selectable marker genes will have a distinctive phenotype for purposes of identification, i.e., they will be distinguishable from non-transformed cells. The characteristic phenotype allows the identification of cells, cell groups, tissues, organs, plant parts, or whole plants containing the construct.

Detection of the marker phenotype makes possible the selection of cells having a second gene to which the marker gene has been linked. This second gene typically comprises a desirable phenotype that is not readily identifiable in transformed cells, but which is present when the plant cell or derivative thereof is grown to maturity, even under conditions wherein the selectable marker phenotype itself is not apparent.

The use of such a marker for identification of plant cells containing a plastid construct has been described by

Svab et al. (1993, *supra*). In the examples provided below, a bacterial *aadA* gene is expressed as the marker under the regulatory control of chloroplast 5' promoter and 3' transcription termination regions, specifically the regulatory regions of the *psbA* gene (described in Staub et al., *EMBO J.* (1993) 12(2):601-606). Numerous additional promoter regions may also be used to drive expression of the selectable marker gene, including various plastid promoters and bacterial promoters that have been shown to function in plant plastids.

Expression of the *aadA* gene confers resistance to spectinomycin and streptomycin and thus allows for the identification of plant cells expressing this marker. The *aadA* gene product allows for continued growth and greening of cells whose chloroplasts comprise the selectable marker gene product. Cells that do not contain the selectable marker gene product are bleached. Selection for the *aadA* gene marker is thus based on identification of plant cells that are not bleached by the presence of streptomycin, or more preferably spectinomycin, in the plant growth medium.

A number of markers have been developed for use with plant cells, such as resistance to chloramphenicol, the aminoglycoside G418, hygromycin, or the like. Other genes that encode a product involved in chloroplast metabolism may also be used as selectable markers. For example, genes that provide resistance to plant herbicides such as glyphosate, bromoxynil or imidazolinone may find particular use. Such genes have been reported (Stalker et al., *J. Biol. Chem.* (1985) 260:4724-4728 (glyphosate resistant EPSP); Stalker et

al., *J. Biol. Chem.* (1985) 263:6310-6314 (bromoxynil resistant nitrilase gene); and Sathasivan et al., *Nucl. Acids Res.* (1990) 18:2188 (AHAS imidazolinone resistance gene)).

5       Stable transformation of tobacco plastid genomes by particle bombardment is reported (Svab et.al. (1990), *supra*) and Svab et al. (1993), *supra*). The methods described therein may be employed to obtain plants homoplasmic for plastid expression constructs.

10       Generally, bombarded tissue is cultured for approximately 2 days on a cell division-promoting media, after which the plant tissue is transferred to a selective media containing an inhibitory amount of the particular selective agent, as well as the particular hormones and  
15 other substances necessary to obtain regeneration for that particular plant species. Shoots are then subcultured on the same selective media to ensure production and selection of homoplasmic shoots.

Transplastomic tobacco plants are analyzed for a pure  
20 population of transformed plastid genomes (homoplasmic lines). Homoplasmy is verified using Southern analysis employing nucleic acid probes spanning a region of the transgene and chloroplast genome (i.e. the insertion region). Transplastomic plants that are heteroplasmic  
25 (i.e., contain a mixture of plastid genomes containing and lacking the transgene) are characterized by a hybridization pattern of wild type and transgenic bands. Homoplasmic plants show a hybridization pattern lacking the wild type band.

Alternatively, homoplasmy may be verified using the polymerase chain reaction (PCR). PCR primers are utilized that are targeted to amplify from sequences from the insertion region. For example, a pair of primers may be  
5 utilized in a PCR reaction. One primer amplifies from a region in the transgene, whereas the second primer amplifies from a region proximal to the insertion region towards the insertion region. A second PCR reaction is performed using primers designed to amplify the region of insertion.  
10 Transplastomic lines identified as homoplasmic produce the expected size fragment in the first reaction, whereas they do not produce the predicted size fragment in the second reaction.

Where transformation and regeneration methods have been  
15 adapted for a given plant species, either by *Agrobacterium*-mediated transformation, bombardment or some other method, the established techniques may be modified for use in selection and regeneration methods to produce plastid-transformed plants. For example, the methods described  
20 herein for tobacco are readily adaptable to other solanaceous species, such as tomato, petunia and potato.

For transformation of soybean, particle bombardment as well as *Agrobacterium*-mediated nuclear transformation and regeneration protocols have been described (Hinchee et al.  
25 USPN 5,416,011, and Christou et al. USPN 5,015,580). The skilled artisan will recognize that protocols described for soybean transformation may be used

In *Brassica*, *Agrobacterium*-mediated transformation and regeneration protocols generally involve the use of

hypocotyl tissue, a non-green tissue that might contain a low plastid content. Thus, for *Brassica*, preferred target tissues would include microspore-derived hypocotyl or cotyledonary tissues (which are green and thus contain numerous plastids) or leaf tissue explants. Although the regeneration rates from such tissues may be low, positional effects, such as seen with *Agrobacterium*-mediated transformation, are not expected, thus it would not be necessary to screen numerous successfully transformed plants in order to obtain a desired phenotype.

For cotton, transformation of *Gossypium hirsutum* L. cotyledons by co-cultivation with *Agrobacterium tumefaciens* has been described by Firoozabady et al., *Plant Mol. Bio.* (1987) 10:105-116 and Umbeck et al., *Bio/Technology* (1987) 5:263-266. Again, as for *Brassica*, this tissue may contain insufficient plastid content for chloroplast transformation. Thus, as for *Brassica*, an alternative method for transformation and regeneration of alternative target tissue containing chloroplasts may be desirable, for instance targeting green embryogenic tissue.

Other plant species may be similarly transformed using related techniques. Alternatively, microprojectile bombardment methods, such as described by Klein et al. (*Bio/Technology* 10:286-291) may also be used to obtain nuclear transformed plants comprising the viral single subunit RNA polymerase expression constructs described herein. Cotton transformation by particle bombardment is reported in WO 92/15675, published September 17, 1992. Suitable plants for the practice of the present invention

include, but are not limited to, soybean, cotton, alfalfa, oil seed rape, flax, tomato, sugar beet, sunflower, potato, tobacco, *maize*, wheat, rice and lettuce.

The vectors for use in plastid transformation preferably include means for providing a stable transfer of the plastid expression construct and selectable marker construct into the plastid genome. This is most conveniently provided by regions of homology to the target plastid genome. The regions of homology flank the construct to be transferred and provide for transfer to the plastid genome by homologous recombination, via a double crossover into the genome. The complete DNA sequence of the plastid genome of tobacco has been reported (Shinozaki et al., *EMBO J.* (1986) 5:2043-2049). Complete DNA sequences of the plastid genomes from liverwort (Ohyama et al., *Nature* (1986) 322:572-574) and rice (Hiratsuka et al., *Mol. Gen. Genet.* (1989) 217:185-194) have also been reported.

Where the regions of homology are present in the inverted repeat regions of the plastid genome (known as IRA and IRB), two copies of the transgene are expected per transformed plastid. Where the regions of homology are present outside the inverted repeat regions of the plastid genome, one copy of the transgene is expected per transformed plastid. The regions of homology within the plastid genome are approximately 1kb in size. Smaller regions of homology may also be used, and as little as 100 bp can provide for homologous recombination into the plastid genome. However, the frequency of recombination and thus the frequency of obtaining plants having transformed



plastids decreases with decreasing size of the homology regions.

Examples of constructs having regions of homology the plastid genome are described in Svab et.al. (1990 *supra*),  
5 Svab et al. (1993 *supra*) and Zoubenko et al. (*Nuc Acid Res* (1994) 22(19):3819-3824).

Expression constructs employing the downstream box sequences as well as the translation initiation regions of the present invention find use in directing the expression  
10 of DNA sequences encoding genes involved in a wide variety of plant genetic engineering applications. Such genes may encode for proteins involved in agronomic traits (input traits) such as herbicide tolerance and disease resistance, or quality traits (output traits) such as fatty acid  
15 composition modification and carotenoid production. Furthermore, DNA sequences encoding for proteins for the production of human biologics in a plant cell plastid also find use in the expression constructs of the present invention.

20 As described in more detail in the examples below, constructs are prepared to direct the expression of a DNA sequence coding for an enzyme involved in the degradation of cellulose. The constructs described comprise DNA sequences encoding a thermophilic E1 cellulase from *Acidothermus*  
25 *cellulolyticus* (U.S. Patent Number 5,536,655, the entirety of which is incorporated herein by reference), utilizing the T7 polymerase expression system (described in U.S. Patent Number 5,576,198, the entirety of which is incorporated herein by reference). The constructs further comprise

downstream box sequences described herein. Other sequences encoding for genes involved in degrading cellulose also find use in constructs employing the DB sequences of the present invention. For example, DNA sequences encoding for other  
5 polysaccharide hydrolyzing enzymes such as those from *Thermomonospora fusca* (See, e.g., Wilson (1992) *Crit. Rev. Biotechnol.* 12:45-63) may be used in the expression constructs of the present invention.

The skilled artisan will recognize that other DNA  
10 sequences find use in the constructs of the present invention.

For example, the expression constructs of the present invention allow for the high level expression of agronomically important traits such as herbicide and stress  
15 tolerance from a plant cell plastid. DNA sequences encoding for proteins involved in herbicide tolerance are known in the art, and include, but are not limited to, DNA sequences encoding for 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS, described in U.S. Patent Numbers 5,627,061 and  
20 5,633,435, Padgett et al. (1996) *Herbicide Resistant Crops*, Lewis Publishers, 53-85, and in Penaloza-Vazquez, et al. (1995) *Plant Cell Reports* 14:482-487) and *aroA* (U.S. Patent Number 5,094,945) for glyphosate tolerance, bromoxynil nitrilase (Bxn) for bromoxynil tolerance (U.S.  
25 Patent Number 4,810,648), phytoene desaturase (*crtI* (Misawa et al. (1993) *Plant Journal* 4:833-840, and (1994) *Plant Jour.* 6:481-489) for tolerance to norflurazon, acetohydroxyacid synthase (AHAS (Sathasiivan et al. (1990) *Nucl. Acids Res.*

18:2188-2193)) and the *bar* gene for tolerance to glufosinate (DeBlock, et al. (1987) *EMBO J.* 6:2513-2519.

It should be noted that the expression constructs of the present invention may also include sequences encoding  
5 genes involved in other stress tolerance genes, for example insect or disease resistance/tolerance genes. Such insect tolerance genes are known in the art, for example, the *Bacillus thuringiensis cry1Ac* protein.

In addition, the expression constructs also find use in  
10 directing the production of human biological proteins (pharmaceutical proteins) from the plant plastid. Nucleic acid sequences encoding for the human growth hormone (hGH) may be employed in the plastid expression constructs of the present invention.

15 Another example of utilizing the expression constructs of the present invention for the production of human biological proteins is the production of aprotinin.

Other sequences that may find use in the production of human biologics include sequences encoding for insulin or  
20 insulin precursors. The skilled artisan will recognize that many nucleotide sequences encoding for human biologics may be employed in the constructs of the present invention to direct their expression from a plant plastid such as those described in Goodman and Gelman (1990) *Pharmacological Basis*  
25 *of Therapeutics*, Pergaman Press, 8<sup>th</sup> Edition, Sections 14 and 15.

Furthermore, the downstream boxes as well as the synthetic translation initiation region of the present invention may also be employed in expression constructs

utilizing the T7 polymerase expression system. Such expression systems are known in the art and are described in U.S. Patent Number 5,576,198, the entirety of which is incorporated herein by reference.

- 5 Other sequences may find use in the expression constructs employing the T7 polymerase expression system, such constructs may allow for the regulation of the sequences in a specific tissue or developmental stage.

Preferential expression in seed plastids is desirable  
10 where modification of fatty acid biosynthesis pathways, modification of storage proteins, or introduction of a new pathway in seeds, such as the bacterial polyhydroxybutyrate (PHB) pathway is desired. Examples of plant functional promoters that may be useful for such applications include  
15 those from plant storage protein genes or from genes involved in fatty acid biosynthesis in oilseeds. Examples of such promoters include the 5' regulatory regions from such genes as napin (Kridl et al. (1991) *Seed Sci. Res.* 1:209:219), phaseolin, zein, soybean trypsin inhibitor, ACP,  
20 stearoyl-ACP desaturase and oleosin. Seed specific gene regulation is discussed in EP 0 255 378 (2/3/88).

Preferential expression in fruit plastids is desirable, for example, where modification of fruit color or flavor development pathways is desired, or where one wishes to  
25 alter the carbohydrate content in plant fruits. Examples of plant functional promoters that may be useful for such applications include those from such fruit related genes as polygalacturonase (Bird et al. (1988) *Plant Mol. Biol.* 11:651-662), the E-8 gene from tomato (Deikman et al. (1988)

EMBO J. 7:3315-3320; DellaPenna et al. (1989) *Plant Cell* 1:53-63), tomato 2A11 (Pear et al. (1989) *Plant Mol. Biol.* 13:639-651), or ovary-specific promoter regions such as described in WO 91/01324 (2/7/91).

- 5        Modification to starch synthesis pathways in major starch storage tissues such as potato tubers or corn seeds may also be accomplished by the plastid expression methods described herein. In such cases, promoters from patatin (Twell et al. (1987) *Plant Mol. Biol.* 9:365-375), zein or  
10 plant starch synthase (Visser et al. (1989) *Plant Sci.* 64:185-192) may be particularly useful for nuclear expression of a T7 polymerase protein.

- Other plant tissues that can be selectively modified by the constructs and methods described herein include plant  
15 floral tissues and such specialized tissues as cotton boll fibers. Promoters that may be used for nuclear expression of a T7 RNA polymerase selectively in floral tissues include those from a chalcone synthase gene, such as the CHS gene A from petunia (Koes et al. (1986) *Nucl. Acids Res.* 14:5229-  
20 5239). For floral color modification, expression or antisense control of anthocyanin or flavonoid type pigments in floral tissues is desired. For a review of plant flower color gene manipulation, see van Tunen et al. (in *Plant Biotechnology Series, Volume 2* (1990) Developmental  
25 Regulation of Plant Gene Expression, D. Grierson ed.). For fiber tissue modification, use of a tomato p7Z gene promoter for expression in cotton fiber tissues is described in US Patent No 5530185. Other promoters useful for expression in cotton fiber cells have also been reported (Crow et al.,

Proc. Natl. Acad. Sci. USA (1992) 89:5769-5773). Desirable cotton fiber modifications include improvement of strength or texture. Examples of genes that may be used for such improvements included PHB biosynthesis genes, cellulose  
5 synthase genes (Saxena et al. (1990) *Plant Mol. Biol.* 15:673-683) and fungal chitin synthase genes (Bowen et al. (Proc. Nat. Acad. Sci. (1992) 89:519-523).

Another example of useful phenotypic modification to cotton fibers is the production of colored cotton, for  
10 example having blue or black colored fibers. Methods for nuclear expression of encoding sequences for pigment production in cotton fibers is described in US Patent No. 5530185. For example, for melanin production, two protein encoding sequences from *Streptomyces* (Bernan et al. (1985)  
15 *Gene* 37:101-110) may be used. The mRNAs for the two proteins, the ORF438 product and tyrosinase, are transcribed from a single promoter *Streptomyces*. By the instant invention, the operon may be adapted for expression from a regulatable promoter region and used for plastid  
20 transformation. Similarly, two or more gene products may be required for production of indigo in transformed host cells. Where the responsible enzyme is a monooxygenase, as described in unexamined Japanese Patent Application Kokai 2-119777 (Suzuki et al.), two gene products are required.  
25 Where the enzyme is a dioxygenase, such as the naphthalene dioxygenase (ND) described in Kurkela et al. (*Gene* (1988) 73:355-362), three gene products are required. Use of a tryptophanase encoding sequence may also be desired to increase the amount of indole available for conversion to

indigo. Sources of tryptophanase gene sequences include *E. coli* (see, for example Deeley et al. (1982) *J. Bacteriol.* 151:942-951).

In a further embodiment, the instant application provides for methods to preferentially modify plastid pathways that are present in green tissues or that are preferentially present in actively growing plant tissues, such as meristematic regions. In particular, use of light inducible promoters, such as those from SSU or chlorophyll A/B binding protein genes, for expression of a RNA polymerase protein in green tissues is considered. For actively growing plant tissues, a promoter from an EF-1 $\alpha$  gene may be used. See, for example, USPN 5,177,011 (Shewmaker et al.) issued 1/5/93. Such green tissue or meristematic promoters may find use in conjunction with plastid expression constructs for modifications that provide herbicide tolerance, for example to glyphosate, bromoxynil or imidazolinone herbicides using resistance genes such as described in Stalker et al. (*J. Biol. Chem.* (1985) 260:4724-4728), Stalker et al. (*J. Biol. Chem.* (1985) 263:6310-6314), and Sathasivan et al. (*Nucl. Acids Res.* (1990) 18:2188). In addition, the plastids of such tissues are desirable targets for modifications to provide increased photosynthetic capacity or to provide mechanisms for disease or stress resistance.

Furthermore, the prokaryotic DB sequences used in the constructs of the present invention may also find use in the identification of DB sequences from eukaryotic sources. Some plastid genes do not contain Shine-Dalgarno sequences.

Because downstream box sequences have been found to initiate translation alone (Sprengart et al. (1996) *supra*), these genes may contain downstream box-like sequences that control translation initiation. Thus, the DB sequences of the  
5 present invention may be used to identify plastidial DB sequences through sequence database searches.

The invention now being generally described, it will be more readily understood by reference to the following examples, which are included for purposes of illustration  
10 only and are not intended to limit the present invention.

### EXAMPLES

#### Example 1 Expression Constructs

15 Constructs and methods for use in transforming the plastids of higher plants are described in Zoubenko et al. (*Nuc Acid Res* (1994) 22(19):3819-3824), Svab et al. (*Proc. Natl. Acad. Sci.* (1990) 87:8526-8530 and *Proc. Natl. Acad. Sci.* (1993) 90:913-917) and Staub et al. (*EMBO J.* (1993)  
20 12:601-606). Constucts and methods for use in transforming plastids of higher plants to express DNA sequences under the control of a nuclearly encoded, plastid targeted T7 polymerase are described in U.S. Patent Number 5,576,198. The complete DNA sequences of the plastid genome of tobacco  
25 are reported by Shinozaki et al. (*EMBO J.* (1986) 5:2043-2049). All plastid DNA references in the following description are to the nucleotide number from tobacco.

A vector was prepared to direct the expression of the *Acidothermus* E1  $\beta$ -1,4-endoglucanase in plant plastids. The



plasmid pMPT4, a pGEM (Clonotech) derivative containing the entire *Acidothermus* E1 cellulase coding sequence (U.S. Patent Number 5,536,655) and flanking regions on a 3.7 kb *Pvu* I genomic DNA fragment, was digested with the

5 restriction endonuclease sites *Sac*II and *Asp*718 to remove the coding sequence for the mature E1 cellulase protein. This fragment was cloned into the same restriction sites of the plasmid pBCSK+ (Stratagene) to create the vector pCGN6063. This plasmid was digested with *Sac*I and *Sac*II and

10 a double-stranded oligonucleotide sequence, 5'-GGAGCTGCGTCACCATGGCGGGA-3' (SEQ ID NO: 7), was inserted to introduce an *Nco*I site-derived ATG translational start codon fused to the 60,000 mol wt mature sequence (minus the endogeneous bacterial signal peptide amino acid sequence) of

15 the E1  $\beta$ -1,4-endoglucanase polypeptide, creating the construct pCGN6067. The E1 gene was excised from pCGN6067 as an *Nco* I to *Asc* I DNA segment and cloned into the T7 promoter expression cassette pCGN5063 to create the construct pCGN6108. This plasmid contains the plastid

20 expression regulatory elements of the T7 bacteriophage promoter operably linked to the mature protein portion of the E1 coding sequence and *psbA* transcription termination region. The chimeric cellulase expression cassette was excised as a *Hind* III to *Not* I DNA fragment and cloned into

25 the tobacco chloroplast homology vector, pCGN6043 in the same restriction sites to create the construct pCGN6115. The homology sequences employed in the vector direct the integration of the E1 cellulase gene and *aadA* marker transgene to the region between the *rbcL* and ORF512

sequences (described in Svab et al., (1993) *supra*). The construct pCGN6115 (Figure 1) was used to transform tobacco plants to direct the transformation to homoplasmy and plastid expression of the E1  $\beta$ -1,4-endoglucanase encoding gene in the plant plastid.

A series of constructs is prepared containing various SD and DB sequences. Polymerase chain reactions (PCR) are employed using various primer combinations to amplify fragments containing the SD and DB sequences. Reactions using oligonucleotide primers and plasmid pCGN5063 in pBluescript + (Stratagene) containing the GUS gene under the control of T7 promoter and T7gene10 leader (similar to pCGN4055 described in U.S. Patent 5,576,198) are performed. The T7 promoter and T7gene10 leader is present as a HindIII/NcoI fragment. The primer pairs for each construct are designed to introduce HindIII and NcoI at the respective ends. The resulting PCR fragments are purified, digested with HindIII/NcoI, and ligated to HindIII/NcoI digested pCGN5063 vector backbone. The forward primer carrying the HindIII site is designed to prime at the T7 promoter region and is common for all the constructs (G10L5' 5'-ACGTAAGCTTCGAAATT AATACGACTCACTATAGGG-3') (SEQ ID NO: 8). The reverse primer containing the NcoI site introduced the downstream box and Shine-Dalgarno sequence variants for the various constructs and are listed in Table 1. The downstream box (DB) variants include a) wildtype gene 10 DB (wt DB), which has 7 bases complementary to the plastidial 16S rRNA(7/15); (b) mutant DB with 15 bases complementary to the plastidial 16S rRNA(m1DB, 15/15); (c) mutant DB with 11

bases complementary to the plastidial 16S rRNA(m2DB, 11/15);  
 and (d) mutant DB with 0 bases (m3DB, 0/15) that potentially  
 can pair with the 15 basepair anti-DB sequence in the  
 tobacco 16S rRNA. The Shine-Dalgarno sequence (SD) variants  
 5 include wildtype SD AAGGAG (wt SD) and mutant SD UUCCUC  
 (mSD). In the next cloning step, the GUS coding region was  
 removed as a NcoI/AscI fragment and replaced by NcoI/AscI E1  
 cellulase fragment from p6067. Sequences representing the  
 various combinations of SD sequences with the initiation  
 10 codons and DB combinations are presented in Figures 2  
 through 7. Base pairing between the wild type T7 gene 10 SD  
 and DB sequences with complementary sequences of the  
 bacterial 16S rRNA and plastidial 16S rRNA is shown in  
 Figure 8.

15

Table 1

CONSTRUCT	FEATURES	REVERSE PCR PRIMER
PCGN6376	wt SD, wt DB (7/15)	SC123: (SEQ ID NO: 9) 5'ACTGCCATGGCCATTGCTGTC CACCAGTCATGCTAGCCATATGT ATATCTCCTTCTTAAAGTTAAAC
PCGN6115	wt SD, Δ DB	
PCGN6377	m SD, wt DB (7/15)	SC125: (SEQ ID NO: 10) 5'ACTGCCATGGCCATTGCTGTC CACCAGTCATGCTAGCCATATGT ATATGAGGAACTTAAAGTTAAAC AAAATTAT
PCGN6365	wt SD, m1 DB (15/15)	SC126: (SEQ ID NO: 11) 5'ACTGCCATGGCCATTGCAAGG CAGGACTAATGATAGCCATATGT ATATCTCCTTCTTAAAGTTAAAC
PCGN6367	m SD, m1 DB (15/15)	SC127: (SEQ ID NO: 12) 5'ACTGCCATGGCCATTGCAAGG CAGGACTAATGATAGCCATATGT ATATGAGGAACTTAAAGTTAAAC

pCGN6368	wt SD, m2 DB (11/15)	SC128: (SEQ ID NO: 13) 5'ACTGCCATGGCCATTTGCTGTC GGCCTGACCACCTAGCCATATGT ATATCTCCTTCTTAAAGTTAAAC
pCGN6369	wt SD, m3 DB (0/15)	SC129: (SEQ ID NO: 14) 5'ACTGCCATGGCCATTTGCTGGG CAGCGGTAGTGCTAGCCATATGT ATATCTCCTTCTTAAAGTTAAAC

---

In the final cloning step, E1 cellulase expression cassette is cut with HindIII/NotI and ligated to the plastid transformation vector backbone of pCGN6154 digested with

5 HindIII/NotI. pCGN6154 is derived from pCGN6049, the plastid transformation vector that targets to the plastid 16S repeat region. pCGN6154 contains the gene cassette for expression of aprotinin as HindIII/NotI fragment, which is removed by gel purification, and the vector backbone was used for the

10 cloning.

Cloning the various SD and DB sequences and E1 cellulase fragments into the 6154 backbone vector results in the production of binary vectors for transformation (Figure 9). The vector pCGN6376 contains the T7 Promoter, wild-type

15 SD, the 7/15 DB, and E1 cellulase. The vector pCGN6364 contains the T7 Promoter, wild-type SD, no DB sequence, and E1 cellulase. The vector pCGN6377 contains the T7 Promoter, mutant SD, the 7/15 DB, and E1 cellulase. The vector

pCGN6365 contains the T7 Promoter, wild-type SD, the 15/15

20 DB, and E1 cellulase. The vector pCGN6367 contains the T7 Promoter, mutant SD, the 15/15 DB, and E1 cellulase. The vector pCGN6368 contains the T7 Promoter, wild-type SD, the 11/15 DB, and E1 cellulase. The vector pCGN6369 contains the T7 Promoter, wild-type SD, the 0/15 DB, and E1 cellulase.

## Example 2 Plant Transformation

### Plastid Transformation

Tobacco plastids are transformed by particle gun  
5 delivery of microprojectiles as described by Svab and Maliga  
(*Proc. Natl. Acad. Sci.* (1993) 90:913-917) and described  
here.

Tobacco plastids are transformed by particle gun  
delivery of microprojectiles. Because integration into the  
10 plastid genome occurs by homologous recombination and the  
target site is between the 16S rRNA and trnV in the repeat  
region of the plastid genome, two copies of the transgene  
are expected per plastid genome (Svab et al. (1993) *supra*).

Tobacco seeds (*N. tabacum* v. Xanthi N/C) homozygous for  
15 pCGN4026 (McBride et al., U.S. Patent Number 5,576,198) T-  
DNA are surface sterilized in a 50% chlorox solution (2.5%  
sodium hypochlorite) for 20 minutes and rinsed 4 times in  
sterile H<sub>2</sub>O. The seeds are then plated aseptically on a  
0.2X MS salts media and allowed to germinate. The seedlings  
20 are grown on agar solidified MS media with 30g/L sucrose  
(Murashige and Skoog (1962) *Physiol. Plant* 15:493-497).

Gold microprojectiles (0.6µm) are coated with DNA,  
such as the T7/E1 cellulase expression construct, pCGN6115  
(Fig. 1), and the coated microprojectiles used to bombard  
25 mature leaves, placed abaxial side up on RMOP media (MS  
salts, 1 mg/L BAP, 0.1 mg/L NAA, 30 g/L sucrose and 0.7%  
phytagar) (Svab et al. (1990) *supra*) using the Bio-Rad PDS  
1000/He bombardment system (Sanford et al. (1991) *Technique*  
3:3-16; Klein et al. (1992) *Bio/Technology* 10:286-291).

Development of transformed plants on RMOP media supplemented with 500 mg/L spectinomycin dihydrochloride and subsequent subcloning on the same selective medium is conducted according to Svab et al. (1990); Svab and Maliga (1993);  
5 supra). Selected plants are rooted in MS media containing 1 mg/L IBA, 500 mg/L spectinomycin dihydrochloride and 0.6% phytagar.

### Example 3 Analysis of Transplastomic Tobacco Plants

10

To confirm homoplasmy by Southern hybridization, total plant cellular DNA is prepared using DNAzol reagent (BRL Lifetechnologies, Gaithersburg, MD) following the manufacturer's protocol. Approximately 12 µg DNA for each  
15 sample is digested with *Bam*HI, electrophoresed through 0.7% agarose, and transferred to Nytran+ (Schleicher and Schuell). The filters are hybridized at 50°C in buffer provided with the DNAzol reagent provided by manufacturer and containing labelled probe prepared using manufacturer's  
20 directions. The hybridization probe was prepared from nucleic acid sequences spanning the integration zone.

To demonstrate that the homoplasmic tobacco lines express the E1  $\beta$ -1,4-endoglucanase, Western blot analysis was performed using total soluble leaf protein. Leaf  
25 protein was extracted as follows: 200mg mature leaf samples were frozen in liquid N<sub>2</sub> and ground in 0.08 mL extraction buffer containing 0.1M NaPO<sub>4</sub>, pH6.8, 0.15M NaCl, 0.01M EDTA, 0.01M DTT, 0.01M thiourea, 0.3% Tween-20, and 0.05% Triton-X100. Protein concentrations were determined by Bradford

assay. Protein samples were combined with an equal volume of 2x Laemmli sample buffer (Laemmli (1970) *Nature* 227:680-685) and boiled prior to loading onto 10% Laemmli gels. Approximately 10µg of total leaf protein was loaded per  
5 lane. Western blot analysis is performed using the ECL plus kit (Amersham).

Results of the Western blot analysis using monoclonal antibodies raised against the *Acidothermus* E1 cellulase demonstrate that the E1 protein is expressed in homoplasmic  
10 6368 lines. The purified protein (control) runs as multiple forms on a denaturing gel, the highest form being 72,000 molecular weight as this form includes the signal peptide for secretion. The mature form of the enzyme is around 60 kd.

15 The E1 cellulase expressed in plant plastids as a 60 kD mature form and can be converted to the 40 kD catalytic domain form, presumably by proteolytic processing *in vivo*. Furthermore, from the results of the Western blot analysis, it can be estimated that protein expression of E1  $\beta$ -1,4-  
20 endoglucanase is approximately 0.5% to 1% of the total soluble plant protein in leaves of transplastomic pCGN6115 tobacco lines. However, the estimated protein expression of E1  $\beta$ -1,4-endoglucanase is approximately 4% to 6% of the total soluble plant protein in leaves of transplastomic  
25 pCGN6368 tobacco lines.

Crude total soluble leaf protein from homoplasmic pCGN6115 tobacco lines expressing E1 cellulase are further analyzed for cellulase activity. Because *Acidothermus* E1 cellulase  $V_{max}$  is near maximal approaching 80°C, experiments

are carried out at 55°C and 80°C. Protein extracts are tested in reactions to measure the hydrolysis of the fluorogenic substrate 4-methylumbelliferyl- $\beta$ -D-cellobioside (MUC) as described in Laymon et al. (1996) *Applied Biochem.*

- 5 *Biotechnol.* 57/58:389-397. Increased cellulose degradation at higher temperatures using the protein extracts of leaves transformed to express the E1 cellulase confirms that the cellulase protein expressed from the plant cell plastid demonstrates wild type activity.

10

The above results demonstrate that the downstream box sequences of the present invention provide a method for increasing expression of DNA sequences in a plant cell plastid.

- 15 All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each
- 20 individual publication or patent application was specifically and individually indicated to be incorporated by reference.

- Although the foregoing invention has been described in some detail by way of illustration and example for purposes
- 25 of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claim.



## CLAIMS

I claim:

- 5           1. In a method for the transformation of a plant plastid comprising introducing into a plant plastid a construct having a promoter region functional in a plant plastid linked to a DNA sequence of interest having a translation initiation site and a coding region, the  
10 improvement comprising:  
operably coupling a downstream box sequence to the coding region side of said translation initiation site, wherein said downstream box sequence has a region complementary to a region of ribosomal RNA that associates  
15 with message RNA during initiation of translation.
- 20           2. The method according to Claim 1 wherein said downstream box sequence is complementary to a region in the plastid ribosomal RNA.
3. The method according to Claim 1, wherein said downstream box sequence begins from about 1 to about 35 bases from said initiation site.
- 25           4. The method according to Claim 1, wherein said construct further comprises a Shine-Dalgarno sequence associated with and upstream from said translation initiation site.
- 30           5. The method according to Claim 1, wherein initiation site is from about 1 to about 35 bases downstream from said Shine-Dalgarno sequence.
- 35           6. The method according to Claim 1, wherein said downstream box sequence is complementary to between about 7 and about 15 bases of the nucleotide sequence consisting of AGUGAUCGGGACGGC (SEQ ID NO:15), including bases complementary through G-U base-pairing.

7. The method according to Claim 5, wherein said downstream box sequence is the sequence GCAUGACUGGUGGAC (SEQ ID NO:16).

8. The method according to Claim 5, wherein said downstream box sequence is a synthetic sequence.

9. The method according to Claim 7, wherein said synthetic sequence is GCACUACCGCUGCCC (SEQ ID NO:17).

10. The method according to Claim 7, wherein said synthetic sequence is UCAUUAGUCCUGCCU (SEQ ID NO:18).

11. The method according to Claim 1, wherein said construct further comprises DNA regions of homology to the genome of said plastid, wherein said regions of homology flank said promoter region and said DNA sequence of interest.

12. A plant cell comprising a plant plastid produced according to the method of Claim 1.

13. A construct for transforming plastids of a plant cell, said construct comprising the following as operably joined components in the 5' to 3' direction of transcription:

- (a) a translational initiation site;
- (b) a DNA sequence of interest having a coding region;

and

- (c) a transcription termination region;

wherein said DNA sequence of interest further comprises a downstream box sequence operably coupled to the coding region side of said initiation site, wherein said downstream box sequence comprises a region complementary to a region of ribosomal RNA that associates with message RNA during initiation of translation.

14. The construct according to Claim 13 wherein said downstream box sequence is complementary to a region in the plastid ribosomal RNA.

5 15. The construct according to Claim 1, wherein said downstream box sequence begins from about 1 to about 35 bases from said initiation site.

10 16. The construct according to Claim 1, wherein said construct further comprises a Shine-Dalgarno sequence associated with and upstream from said translation initiation site.

15 17. The construct according to Claim 1, wherein initiation site is from about 1 to about 35 bases downstream from said Shine-Dalgarno sequence.

20 18. The construct according to Claim 1, wherein said downstream box sequence is complementary to between about 7 and about 15 bases of the nucleotide sequence consisting of AGUGAUCGGGACGGC (SEQ ID NO:15), including bases complementary through G-U base-pairing.

25 19. The construct according to Claim 5, wherein said downstream box sequence is the sequence GCAUGACUGGUGGAC (SEQ ID NO:16).

30 20. The construct according to Claim 5, wherein said downstream box sequence is a synthetic sequence.

21. The construct according to Claim 7, wherein said synthetic sequence is GCACUACCGCUGCCC (SEQ ID NO:17).

35 22. The construct according to Claim 7, wherein said synthetic sequence is UCAUAGUCCUGCCU (SEQ ID NO:18).

23. The construct according to Claim 1, wherein said construct further comprises DNA regions of homology to the genome of said plastid, wherein said regions of homology

flank said promoter region and said DNA sequence of interest.

24. A plant cell comprising a plant plastid produced  
5 according to the construct of Claim 1.

25. A construct for transforming plastids of a plant cell, said construct comprising the following as operably joined components in the 5' to 3' direction of  
10 transcription:  
    (a) a translational initiation site;  
    (b) a DNA sequence of interest having a coding region;  
    and  
    (c) a transcription termination region;  
15 wherein said DNA sequence of interest further comprises a downstream box sequence to the coding region side of said translation initiation site, wherein said downstream box sequence is within about 30 bases of said translation initiation site, and wherein said downstream box sequence  
20 produces an RNA complementary to between about 7 to about 15 bases of the nucleotide sequence consisting of AGUGAUCGGGACGGC, including bases complementary through G-U base-pairing.

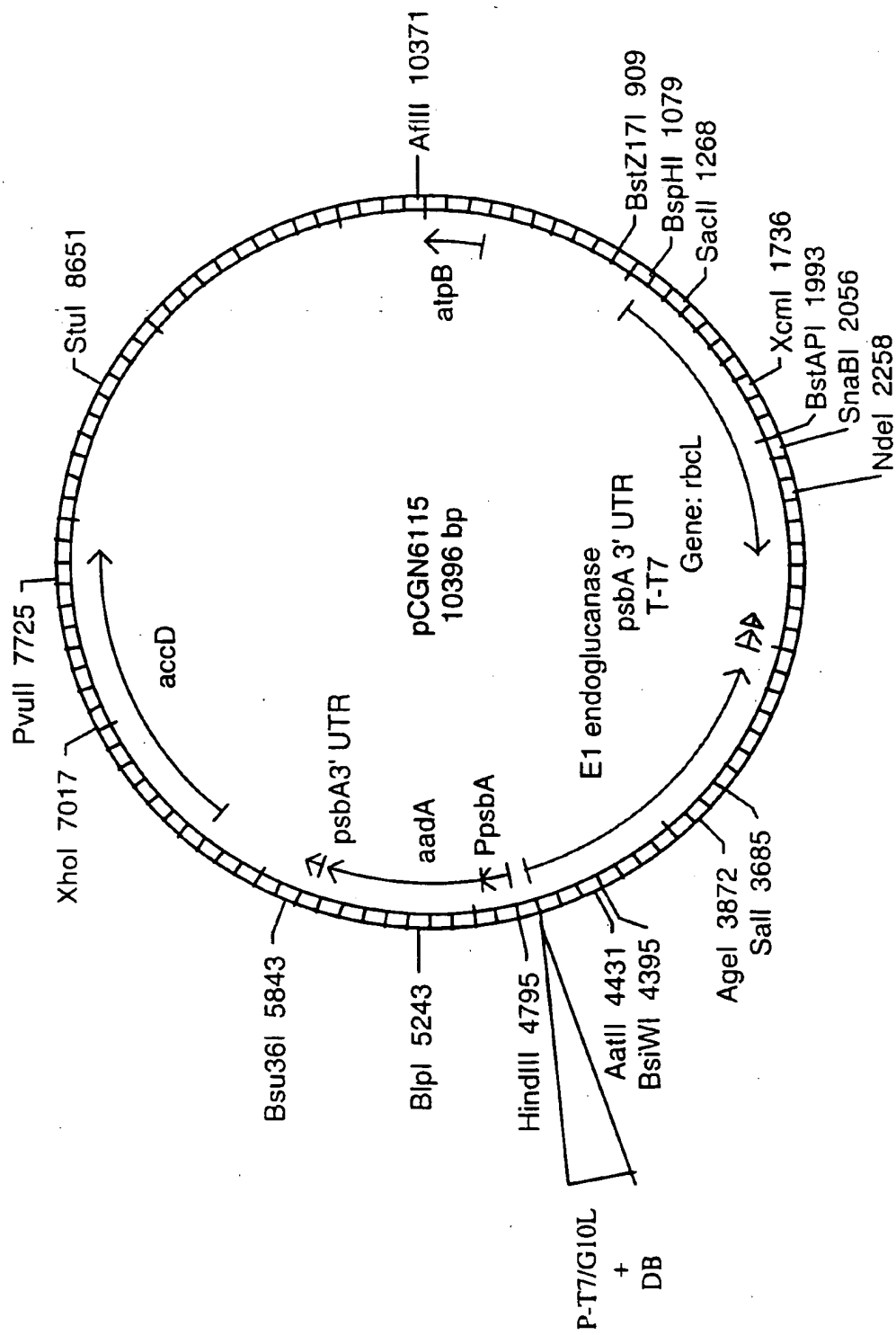


FIGURE 1

5'ACTGCCATGGCCATTTGCTGTCCA<sup>Δ</sup>CAGTCATGCTAGCCATATGTATATCTC  
CTTCTTAAAGTTAAAC

Figure 2

5'ACTGCCATGGCCATTTGCTGTCCACCAGTCATGCTAGCCATATGTATATGAG  
GAACTTAAAGTTAAACAAAATTAT

Figure 3

5'ACTGCCATGGCCATTTGCAAGGCAGGACTAATGATAGCCATATGTATATCT  
CCTTCTTAAAGTTAAAC

Figure 4

5'ACTGCCATGGCCATTTGCAAGGCAGGACTAATGATAGCCATATGTATATGA  
GAACTTAAAGTTAAAC

Figure 5

5'ACTGCCATGGCCATTTGCTGTCGGCCTGACCACCTAGCCATATGTATATCTC  
CTTCTTAAAGTTAAAC

Figure 6

5'ACTGCCATGGCCATTTGCTGGGCAGCGGTAGTGCTAGCCATATGTATATCT  
CCTTCTTAAAGTTAAAC

Figure 7

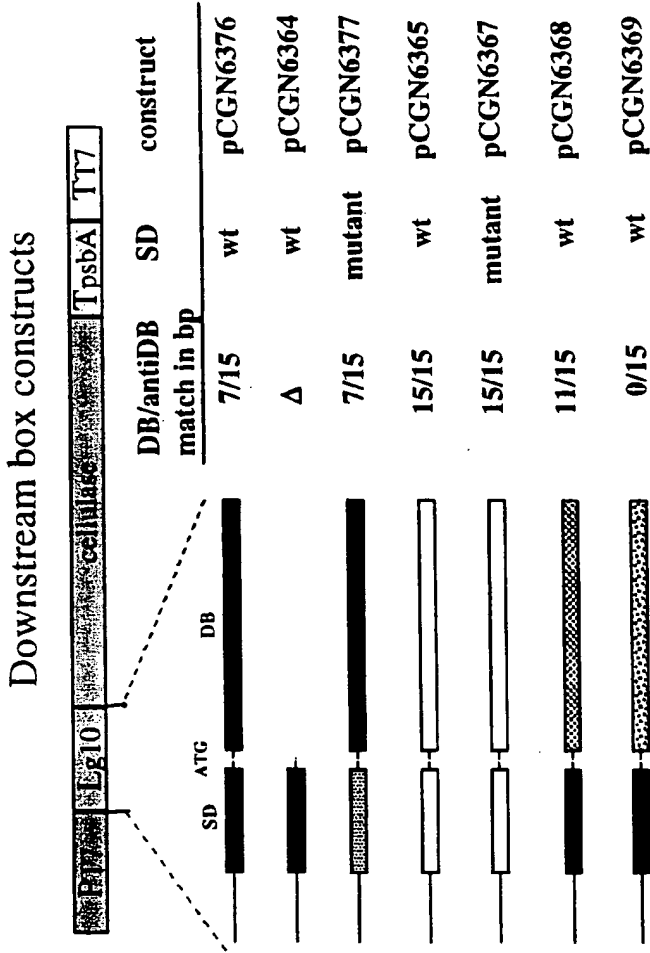


FIGURE 8

T7 gene10 mRNA leader-16S rRNA interactions  
in bacteria and plastid

---

“SD”                      “downstream box”

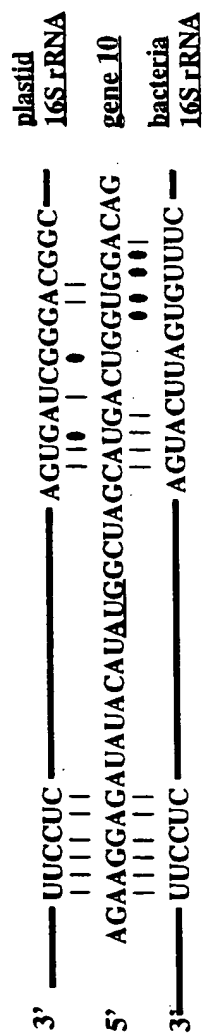


FIGURE 9



## SEQUENCE LISTING

&lt;110&gt; Chaudhuri, Sumita

&lt;120&gt; Enhancer elements for increased translation in plant plastids

&lt;130&gt; 38-21(15340)

&lt;150&gt; US 60/156071

&lt;151&gt; 1999-09-24

&lt;160&gt; 18

&lt;170&gt; PatentIn version 3.0

&lt;210&gt; 1

&lt;211&gt; 68

&lt;212&gt; DNA

&lt;213&gt; Bacteriophage T7

&lt;400&gt; 1

actgccatgg ccatttgctg tccaccagtc atgctagcca tatgtatatc tccttcttaa 60

agttaaac 68

&lt;210&gt; 2

&lt;211&gt; 76

&lt;212&gt; DNA

&lt;213&gt; synthetic construct

&lt;400&gt; 2

actgccatgg ccatttgctg tccaccagtc atgctagcca tatgtatatg aggaacttaa 60

agttaaacaa aattat 76

&lt;210&gt; 3

&lt;211&gt; 68

&lt;212&gt; DNA

&lt;213&gt; synthetic construct

&lt;400&gt; 3

actgccatgg ccatttgcaa ggcaggacta atgatagcca tatgtatatc tccttcttaa 60

agttaaac 68

&lt;210&gt; 4

&lt;211&gt; 68

&lt;212&gt; DNA

&lt;213&gt; synthetic construct

&lt;400&gt; 4

actgccatgg ccatttgcaa ggcaggacta atgatagcca tatgtatatg aggaacttaa 60

agttaaac 68

&lt;210&gt; 5

&lt;211&gt; 68

&lt;212&gt; DNA

&lt;213&gt; synthetic construct

actgccatgg ccatttgctg tcggcctgac cacctagcca tatgtatata tccttcttaa 60  
 agttaaac 68

<210> 6  
 <211> 68  
 <212> DNA  
 <213> synthetic construct

<400> 6  
 actgccatgg ccatttgctg ggcagcggta gtgctagcca tatgtatata tccttcttaa 60  
 agttaaac 68

<210> 7  
 <211> 23  
 <212> DNA  
 <213> synthetic construct

<400> 7  
 ggagctgcgt caccatggcg gga 23

<210> 8  
 <211> 36  
 <212> DNA  
 <213> synthetic construct

<400> 8  
 acgtaagctt cgaaattaat acgactcact ataggg 36

<210> 9  
 <211> 68  
 <212> DNA  
 <213> synthetic construct

<400> 9  
 actgccatgg ccatttgctg tccaccagtc atgctagcca tatgtatata tccttcttaa 60  
 agttaaac 68

<210> 10  
 <211> 76  
 <212> DNA  
 <213> synthetic construct

<400> 10  
 actgccatgg ccatttgctg tccaccagtc atgctagcca tatgtatata aggaacttaa 60  
 agttaaaca aattat 76

<210> 11  
 <211> 68  
 <212> DNA  
 <213> synthetic construct

<400> 11  
 actgccatgg ccatttgcaa ggcaggacta atgatagcca tatgtatata tccttcttaa 60

agttaaac 68

<210> 12  
<211> 68  
<212> DNA  
<213> synthetic construct

<400> 12  
actgccatgg ccatttgcaa ggcaggacta atgatagcca tatgtatatg aggaacttaa 60  
agttaaac 68

<210> 13  
<211> 68  
<212> DNA  
<213> synthetic construct

<400> 13  
actgccatgg ccatttgctg tcggcctgac cacctagcca tatgtatata tccttcttaa 60  
agttaaac 68

<210> 14  
<211> 68  
<212> DNA  
<213> synthetic construct

<400> 14  
actgccatgg ccatttgctg ggcagcggta gtgctagcca tatgtatata tccttcttaa 60  
agttaaac 68

<210> 15  
<211> 15  
<212> RNA  
<213> synthetic construct

<400> 15  
agugaucggg acggc 15

<210> 16  
<211> 15  
<212> RNA  
<213> synthetic construct

<400> 16  
gcaugacugg uggac 15

<210> 17  
<211> 15  
<212> RNA  
<213> synthetic construct

<400> 17  
gcacuaaccgc ugccc 15

WO 01/21782

PCT/US00/26052

<211> 15  
<212> RNA  
<213> synthetic construct

<400> 18  
ucauuagucc ugccu

15

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
29 March 2001 (29.03.2001)

PCT

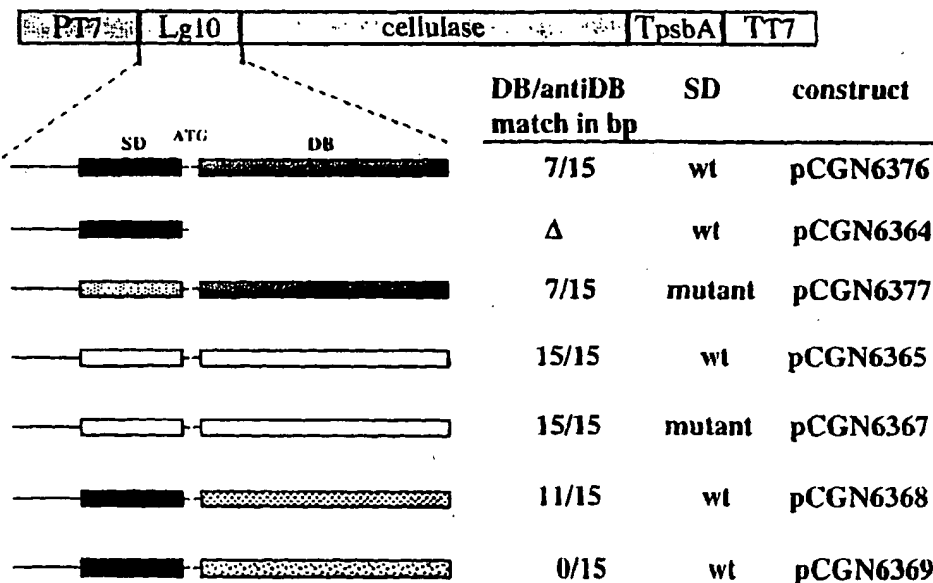
(10) International Publication Number  
WO 01/21782 A3

- (51) International Patent Classification<sup>7</sup>: C12N 15/82, A01H 5/00
- (21) International Application Number: PCT/US00/26052
- (22) International Filing Date:  
22 September 2000 (22.09.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/156,071 24 September 1999 (24.09.1999) US
- (71) Applicant: CALGENE LLC [US/US]; 1920 Fifth Street, Davis, CA 95616 (US).
- (74) Agent: MCBRIDE, Thomas, P.; Monsanto Company, 800 North Lindbergh Boulevard, St. Louis, MO 63167 (US).
- (81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).
- Published:  
— with international search report
- (72) Inventor: CHAUDHURI, Sumita; 1125 Juniper Place, Davis, CA 95616 (US).
- (88) Date of publication of the international search report:  
3 January 2002

[Continued on next page]

(54) Title: ENHANCER ELEMENTS FOR INCREASED TRANSLATION IN PLANT PLASTIDS

## Downstream box constructs



(57) Abstract: Provided are methods for increasing the production of protein in a plant cell by transforming plastids of plant cells with a construct comprising a translation initiation site, DNA sequence of interest and a transcription termination region, and growing plant cells comprising the transformed plastids under conditions wherein the DNA encoding sequence is transcribed in the plastid.



---

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/26052

**A. CLASSIFICATION OF SUBJECT MATTER**  
IPC 7 C12N15/82 A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO 95 16783 A (CALGENE INC ;MCBRIDE KEVIN E (US); STALKER DAVID M (US)) 22 June 1995 (1995-06-22) cited in the application the whole document --- -/--	1-6,8, 11-18, 20,23-25

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

\* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

- \*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- \*X\* document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- \*Y\* document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- \*Z\* document member of the same patent family

Date of the actual completion of the international search

20 April 2001

Date of mailing of the international search report

16/05/2001

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl.  
Fax: (+31-70) 340-3016

Authorized officer

Blanco Urgoiti, B

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/US 00/26052

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	<p>ETCHEGARAY JEAN-PIERRE ET AL:            "Translational enhancement by an element            downstream of the initiation codon in            Escherichia coli."            JOURNAL OF BIOLOGICAL CHEMISTRY,            vol. 274, no. 15,            9 April 1999 (1999-04-09), pages            10079-10085, XP002165654            ISSN: 0021-9258            page 10080; figure 1            page 10081; figure 2            Discussion            page 10083 -page 10084            ---</p>	<p>1-6,8,            11-18,            20,23-25</p>
P,X	<p>WO 00 07431 A (KURODA HIROSHI ;MALIGA PAL            (US); UNIV RUTGERS (US); KHAN MUHAMMAD)            17 February 2000 (2000-02-17)            page 48 -page 50; example 2            figure 2B            ---</p>	<p>1-8,            11-20,            23-25</p>
A	<p>SPRENGART M L ET AL: "THE DOWNSTREAM BOX:            AN EFFICIENT AND INDEPENDENT TRANSLATION            INITIATION SIGNAL IN ESCHERICHIA COLI"            EMBO JOURNAL,GB,OXFORD UNIVERSITY PRESS,            SURREY,            vol. 15, no. 3, 1996, pages 665-674,            XP002922146            ISSN: 0261-4189            cited in the application            the whole document            -----</p>	



# INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No

PCT/US 00/26052

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9516783 A	22-06-1995	US 5576198 A CA 2178645 A EP 0734446 A JP 9510342 T	19-11-1996 22-06-1995 02-10-1996 21-10-1997
WO 0007431 A	17-02-2000	AU 5549099 A	28-02-2000